Meeting review: the Second Meeting on the Critical Assessment of Techniques for Protein Structure Prediction (CASP2), Asilomar, California, December 13–16, 1996

Roland L Dunbrack, Jr, Dietlind L Gerloff, Michael Bower, Xiaowu Chen, Olivier Lichtarge and Fred E Cohen

In most fields of scientific endeavor, the outcomes of important experiments are not always known before the experiments are performed. But in protein structure prediction, algorithms are usually developed and tested in situations where the answers are known. In December 1996, the Second Meeting on the Critical Assessment of Techniques for Protein Structure Prediction (CASP2) was held in Asilomar, California to rectify this situation: protein sequences were provided in advance for which the experimental structure had not yet been published. Over 70 research groups provided bona fide predictions on 42 targets in four categories: comparative or 'homology' modeling, fold recognition or 'threading', *ab initio* structure predictions, and docking predictions. Since the previous CASP meeting in 1994, the role of fold recognition in structure prediction has increased enormously with the largest number of groups participating in this category. In this review, we highlight some of the important developments and give at least a qualitative sense of what kind of methods produced some of the better predictions.

Addresses: Department of Cellular and Molecular Pharmacology, University of California at San Francisco, San Francisco, CA 94143-0450, USA.

Correspondence: Fred E Cohen
e-mail: cohen@cmpharm.ucsf.edu

Electronic identifier: 1359-0278-002-R0027

Folding & Design 01 Apr 1997, 1:R27–R42

Introduction

There are now many approaches to protein structure prediction in situations ranging from homology modeling to *ab initio* protein fold predictions, and the recent Second Meeting on the Critical Assessment of Techniques for Protein Structure Prediction (CASP2) in Asilomar, California was devoted to discussion and comparison of these many approaches. Prediction methods are usually tested on some number of cases where the answer is known, but there is a distinct need for blind predictions as test cases for success. The organizers of the CASP2 meeting gathered the sequences of proteins that were currently under study by NMR spectroscopy or X-ray crystallography from experimental research groups, so that predictions could be made by other groups prior to publication of the structures or deposition of the coordinates in public databanks.

The target sequences were divided into groups depending on the nature of the information known about the protein: (1) comparative modeling, for cases when there was clearly a homologue of known structure in the PDB; (2) fold recognition ('threading'), when there was no such structure in the databank and where the goals were to identify the structure in the PDB with the most similar fold and the sequence alignment between the target and the known structure; (3) *ab initio* structure prediction, again when there was no similar structure in the databank, but where the goal was the prediction of the structure without recourse to known folds; and (4) the prediction of the docking of proteins with ligands, either proteins or small molecules. In the cases of fold recognition and *ab initio* prediction, all of the targets were entered in both categories, since it was not known before the structures were determined experimentally whether or not the fold was similar to a fold in the PDB.

In this report, we review some of the more important issues raised during the meeting. We also review some of the assessment data provided by the organizers and referred to the original submitted predictions for descriptions of methods used. We do not describe the results of all predictors or for all targets, but rather try to give a view of the range of techniques used and a general impression of the results of the assessments and the discussions that occurred for each subgroup. The appointed assessors will provide much more detailed analyses to be published in *Proteins* later this year, and our review of the assessments is essentially qualitative.

It is important to note at the outset that because of the small number of targets in each category and because not all groups submitted predictions for all targets in their categories, the results of the CASP process are not statistically significant measures of predictive success for any particular method. There are now efforts to establish benchmark test sets large enough to give statistically meaningful results, with automated assessments using standardized criteria and software. These include the PROSTAR (J Moult) [1] and FRSVR (D Eisenberg) [2] websites for fold recognition. The issue of benchmarks was discussed at the CASP2 meeting, and it is likely that

benchmarks for sidechain construction and other aspects of comparative modeling as well as docking will be available in the near future.

Nevertheless, the blind nature of the Asilomar meeting and the friendly competition that resulted has been extremely useful in testing and refining tools for protein

Table 1

Prediction teams involved in CASP2 (participant names taken from prediction submissions).

GROUP LEADER and team members

Comparative modeling:

R ABAGYAN, T Cardozo, Y Zhou, S Batalov, M Totrov (New York University)
R BRUCCOLERI (Bristol-Myers Squibb, Princeton, NJ)
F COHEN, M Bower, RL Dunbrack, Jr (University of California, San Francisco)
U EGNER (Schering AG, Berlin, Germany)
K FIDELIS (Lawrence Livermore National Laboratory)
M FORSTER (Molecular Simulations, Inc, San Diego, CA)
B HONIG, A Yang (Columbia University)
C LEE (Stanford University)
J MOULT, R Samudrala, M Braxenthaler, B Milash, J Pedersen, R Luo (University of Maryland)
A ŠALI, R Sanchez (Rockefeller University)
M SAQI (Glaxo Medicines Research Center)
M STERNBERG, PA Bates, RM Jackson (Imperial Cancer Research Fund)
M SUTCLIFFE, S Raza (University of Leicester)
W TAYLOR, A Aszódi (National Institute for Medical Research, London, UK)
G VRIEND (EMBL, Heidelberg)
I WEBER, RW Harrison (Thomas Jefferson University)
P WOLYNES (University of Illinois Urbana–Champaign)

Fold recognition:

R ABAGYAN, S Batalov, T Cardozo, V Maiorov, M Totrov, J Webber, Y Zhou (New York University)
N ALEXANDROV, M Bass, V Solovyev, R Luethy, R Zimmer (Amgen, Inc, GMD)
R ALTMAN, L Wei, J Chang (Stanford University)
G BARTON, RR Copley (Oxford University)
S BRYANT, C Hogue, T Madej, A Marchler-Bauer, H Ohkawa (National Library of Medicine)
A COULSON (University of Edinburgh)
I DUBCHAK (University of California Berkeley)
D EISENBERG, Weiss, Rice, D Fischer (University of California Los Angeles)
A ELOFSSON (University of Stockholm)
A FINKELSTEIN, DS Rykunov (Institute of Protein Research, Moscow)
A GODZIK, L Rychlewski, B Zhang, L Jaroszewski, K Pawlowski, B Reva (Scripps Institute)
U HOBOHM (Roche, Basel)
B HONIG, A Yang, L Xiao (Columbia University)
T HUBBARD, JH Park, A Reinhardt (University of Cambridge)
D JONES (University of Warwick)
K KARPLUS, R Hughey, D Haussler, K Sjolander, C Barrett, M Cline, L Grate, M Hansen, R Karchin, R Rivera, C Tarnas, O Winther (University of California Santa Cruz)
R LATHROP, RG Rogers, TF Smith, JV White (University of California Irvine)
T LENGAUER, H Mevissen, R Thiele, R Zimmer (German National Research Center for Information Technology)
R LUETHY, M Bass (Amgen, Inc)
J MOULT, M Braxenthaler, R Luo, B Milash, J Pedersen, R Samudrala (University of Maryland)
P MUNSON, V DiFrancesco (NIH)
A MURZIN (MRC, Cambridge, UK) F PAZOS, O Olmea, B Rost, A Valencia (University of Madrid, EMBL, Heidelberg)
B ROST (EMBL, Heidelberg)
C SANDER, L Holm, D Haussler, R Hughey, K Karplus (EMBL, Heidelberg, University of California Santa Cruz)
Y SANEJOUAND (Unversité Paul Sabatier, Toulouse, France)
M SIPPL (University of Salzburg)
V SOLOVYEV (Baylor College of Medicine)
M STERNBERG, PA Bates, A Lyall, RB Russell, M Saqi, R Sayle (Imperial Cancer Research Fund, Glaxo Medicines Research Center)
W TAYLOR, R Munro (National Institute for Medical Research, London, UK)
P THOMAS (SmithKline Beecham, King of Prussia, PA)
A TORDA, T Huber, C Dyer, T Lu (Australian National University)

Ab initio predictions

R ABAGYAN, M Totrov, T Cardozo,Y Zhou, J Webber (New York University)
F AVBELJ (Institute of Chemistry, Slovenia)
D BAKER, C Bystroff (University of Washington)
JF BAZAN (DNAX Research Institute, CA)
SA BENNER, M Turcotte, DL Gerloff (University of Florida)
FE COHEN, DL Gerloff, M Joachimiak (University of California, San Francisco),
SA Benner, M Turcotte (University of Florida)
D EISENBERG, S Le Grand (University of California Los Angeles)
AV FINKELSTEIN (Institute of Protein Research, Pushchino Russia)
R GOLDSTEIN (University of Michigan)
T HUBBARD, J Park, A Reinhardt (Centre for Protein Engineering, Cambridge UK)
JAAP Flohil (Institute of Applied Physics, Delft Netherlands)
DT JONES (University of Warwick, UK)
T LENGAUER, F Kaden, B-O Boehmer (German Natl Res Center Information Technology)
GR MARSHALL, S Galaktionov (Washington University Center for Mol Design, MO)
J MOULT, R Samudrala, M Braxenthaler, B Milash, J Pedersen, R Luo (CARB, Rockville MD)
P MUNSON, V Di Francesco, Vasudevan (National Institute of Health, MD)
AG MURZIN, A Bates (MRC, Cambridge UK)
DJ OSGUTHORPE (University of Bath, UK)
GD ROSE (Johns Hopkins University, MD)
B ROST (EMBL, Heidelberg Germany)
BV SHESTOPALOV (Institute of Cytology RAS, Russia)
TF SMITH, L Yu, CG Gaitatzes, JV White (Boston University)
V SOLOVYEV, A Salamov (Baylor College of Medicine, TX)
MJ STERNBERG, PA Bates, RD King, RB Russell (Imperial Cancer Research Fund, London UK)
WR TAYLOR, A Aszódi, REJ Munro (National Inst for Medical Research, London UK)
A VALENCIA, O Olmeo, F Pazos (Protein Design Group, Madrid Spain)

Docking:

R ABAGYAN, M Totrov (New York University)
N BLOM (University of Montreal)
C DELISI, S Vajda, Z Weng (Boston University)
M EISENSTEIN (Weizmann Institute of Science)
R HARRISON, I Weber (Thomas Jefferson University, Philadelphia PA)
T LENGAUER, B Kramer, M Rarey (GMD-SCAI)
T MITCHELL (SmithKline Beecham)
R READ, S Ness, T Hart (University of Alberta)
A REES, P Dauber-Osguthorpe DM Webster (University of Bath)
V SOBOLEV (Weizmann Institute of Science)
M STERNBERG (Imperial Cancer Research Fund)
I VAKSER (Rockefeller University)

structure prediction. Many researchers who submitted predictions had previously developed algorithms and software for particular aspects of structure prediction, but to complete work on the CASP2 targets required them to combine their methods with other techniques developed by other groups. This combining of prediction tools is one very positive outcome of the CASP process.

There were nearly 1000 submitted predictions in the four categories by over 70 research groups. We have listed these groups in Table 1 for each prediction category. We tried to identify group leaders from author lists in the submitted predictions, but this was not always clear. In most cases, these were the principal investigators or laboratory directors. We refer to the groups by these names in capital letters in the rest of this review.

Submission and processing of predictions

Dissemination of target information and submission of predictions was handled entirely via the World Wide Web. All target sequences were assembled on the meeting website [3] and research groups and individuals could register and download target information from the Web. Predictions were submitted by e-mail to a server at Lawrence Livermore National Laboratory which evaluated the format of the prediction, usually a PDB file with records specific to the kind of prediction being submitted, and returned an analysis of the file indicating errors of format or missing items in the submission. Once an acceptable entry was indicated (after several tries in our experience), the file could be submitted by e-mail to the main server.

Much of the assessment was automated by the organizers, who processed the entries and made numerical comparisons with the experimental results, such as root-mean-square deviation (rmsd) calculations and the accuracy of predicted sequence alignments. All of the numerical assessments are available on the Web at the Lawrence Livermore Protein Structure Prediction Center site [4]. These results were sent to the independent assessors, one for each subgroup of predictions (comparative modeling, fold recognition, etc.), along with the actual submitted predictions for each protein. These assessors compared the predictions, assessed the different methods used as described by the predictors and prepared summaries for their subgroup of predictions for the meeting.

At the time of writing this review, not all of the actual submissions are available on the Web, although this should be rectified in the near future. In the numerical assessments, the identities of the author teams are also not all readily available, and these too will be made available by the organizers. This is important, because researchers may have different views of what is important in the analysis of the predictions and should have access to all of the submissions and assessments to explore aspects not covered or emphasized by the official assessors.

Comparative modeling

There were 10 targets in the comparative modeling section, two of which were not solved in time for the meeting. However, only six of the targets were attempted by more than one research group, so we review the assessments of these six (T0001, T0003, T0009, T0017, T0024 and T0028). The numerical assessments were performed by Adam Zemla, Ceslovas Venclovas, and Krzysztof Fidelis (Lawrence Livermore National Laboratory), and critical assessment was provided by Janet Thornton (University College London). The targets whose structures were solved in time for the meeting are listed in Table 2 with the resolution of the parent and target structures and their sequence identities. The resolutions of the targets varied greatly from 1.6 to 3.6 Å and the parent–target sequence identities varied from 10 to 85%. As described in many previous studies, both of these parameters have large effects on our ability to make successful models.

The process of comparative modeling (or more commonly 'homology modeling') is usually broken down into a number of steps: alignment of the target sequence to the parent sequence(s) and structure(s); identification and building of the core residues; building of loops and insertions and deletions; and building of sidechains. The numerical assessments provided measures of each of these steps. Each step in the process has serious consequences for the steps following it, since errors at one step will inevitably cause problems in the later steps.

Alignment

The sequence alignment methods used included programs from commercial modeling packages (ICM [5], Tripos' COMPOSER and MSI's MODELLER [6]), other available programs based on the Smith–Waterman, Needleman–Wunsch and hidden Markov methods (GCG's BESTFIT, FASTA and GAP, CLUSTALW, and AMPS), and in-house programs for sequence alignment. Most of the predictors indicated in their submissions that manual adjustments to the automatic alignments were made, reflecting the experience of some predictors in the CASP1 meeting in 1994. As expected from sequence identities spanning such a large range, some alignments were trivial and some were quite difficult. One important point that arose during the discussion at the meeting was that high sequence identity in a short segment of residues can be very misleading. Sequence/sequence alignments almost always have higher identity than structure/structure alignments, since the function optimized in the former is the sequence identity itself, while in the latter case it is the structural overlap that is optimized.

Table 2

Comparative modeling targets.

Target ID	No. residues	Resolution	Name of protein	Species	Parent(s)	Resolution	Parent lengths	% Seq. ID
T0001	162	2.6 Å	Dihydrofolate reductase	*Haloferax volcanii*	1DYI-A	1.9 Å	171	34
T0002	514	n.a.	Threonine deaminase	*Escherichia coli*	1WSY-B	2.5 Å	397	13
T0003	154	2.5 Å	Phosphotransferase IIA, polyribonucleotide domain	*Mycoplasma capricolum*	1GPR 1F3G	1.9 Å 2.1 Å	158 150	44 36
T0009	109	1.6 Å	Cucumber stellacyanin	*Cucumis sativus*	2CBP	2.5 Å	86	27
T0017	217		Glutathione transferase	*Rattus rattus* liver	2GST-A 1HNA	1.8 Å 1.85 Å	217 217	85 76
T0024	158	2.0 Å	UBC9	*Mus musculus*/*Homo sapiens*	1AAK 2UCE	2.4 Å 2.7 Å	151 148	37 33
T0027	359	1.9 Å	Pectin lyase A	*Aspergillus niger*	2PEC-A 1PCL-A	2.2 Å 2.2 Å	353 355	11 10
T0028	371	3.6 Å	Endoglucanase I (catalytic domain)	*Trichoderma reesei*	1CEL-A	1.8 Å	434	41

n.a., not available.

Core building

There are different definitions of 'core' residues, but in this context the core was defined as those residues in the target that have only a small deviation from residues in the template structure, as defined by the structure/structure alignment program SSAP [7]. The assessors used a Cα–Cα distance of less than 3.0 Å between residues after superposition of the structures to define the core. With this definition, from 67 to 92% of residues in the target proteins were defined as core (depending on the protein). These core regions usually comprised the hydrophobic core and conserved secondary structures, but in some cases also included loops whose lengths and positions were conserved between parent and target. As noted by Janet Thornton in her review of the prediction results, the parent structures without any structural adjustments were generally the best models of the core regions. The difficulty lies in how much to copy from the parent in defining the 'core regions'.

Further efforts to alter the parent structure via Monte Carlo or molecular mechanics simulation or minimization served to move the core structure away from both the target and parent structure. Most significant problems for the target cores were caused by sequence alignment problems, causing very large parent/target rmsds (up to 15 Å in some cases).

Loop modeling

Loops or non-core regions were defined as segments of three or more residues that were not part of the core as defined above. Modeling of these segments was performed either with database methods that attempt to find a suitable model of a segment from the PDB and constraints on the length and end-to-end distance for the segment, or *ab initio* methods that use simulation or some other method of construction under the constraints of the protein structure. Database methods are more common and were used by the groups using MODELLER [6] (FORSTER, ŠALI, SUTCLIFFE and WOLYNES) and by other groups including ABAGYAN (ICM [5]), EGNER (INSIGHT and COMPOSER), SAQI (MSI's QUANTA), STERNBERG, and VRIEND (WHAT IF [8]). *Ab initio* techniques were used by the groups of BRUCCOLERI (CONGEN [9]), COHEN (DRAWBRIDGE), FIDELIS (LOOP BUILDER [10]), HONIG, MOULT [10], TAYLOR (DRAGON [11]), and WEBER [12].

The structures of these loops can be evaluated either in terms of global or local structure similarities. In global comparisons, the complete predicted and experimental target structures were aligned and the rmsds for loops calculated. For local comparisons, the loop regions themselves were aligned structurally before rmsds were calculated. In most cases, the local structural alignments were much better than the global alignments, since many loops have very similar structure between homologous proteins, but may move via a hinge-like mechanism between one structure and another. Again, most of the deviations in the models were greater than those between the parent and target structures, indicating that many computational methods are not succeeding at moving the model from the parent to the target.

Sidechain addition

Once a full model of the backbone is created, sidechains must be placed, and there are numerous algorithms for doing so. Most of these use a rotamer approximation where a small number of conformations for each sidechain are built onto the model backbone and checked for steric conflicts.

These rotamers can be from a backbone-independent rotamer library of a type similar to that of Ponder and Richards [13] or of a variety of context-dependent rotamer libraries. Backbone-independent rotamer libraries were used in predictions made by several groups (ŠALI, SUTCLIFFE and WOLYNES) using the program MODELLER [6] and those groups (EGNER and FORSTER) using the INSIGHT package or the ICM method of Abagyan [5,14]. Context-dependent rotamers can be of several types. Rotamer libraries based on secondary structure were used by STERNBERG [15] and in the SYBYL package used by EGNER [16]. Rotamers chosen by reference to segments of 5–7 residues about the central residue whose rotamer is being chosen were used by LEE [17] and by VRIEND using the program WHAT IF [18] The backbone-dependent rotamer library of Dunbrack [19–21] was used by the COHEN group [22,23] and by those groups using MSI's QUANTA (FORSTER and TAYLOR). MOULT and FIDELIS used a library described as dependent on the local backbone conformation.

Overall results

We looked at the results for the six targets with multiple predictions to see if there were some groups whose methods seemed to produce better models. We considered the following criteria in our review of the assessments of the submissions: the number of alignment errors; the mainchain rmsd scores for the entire structure and the core regions; the global mainchain rmsd scores of the loop regions; the percentage of ϕ and ψ values predicted within 30° of the target values; and the percentage of χ_1 values predicted within 30° of their target values.

On the basis of these criteria, it was clear that three groups consistently created models that were more accurate than most of the other models submitted for the same targets: in particular, the MOULT group (targets T0009, T0024 and T0028), the WOLYNES group (T0001, T0017 and T0003), and the ŠALI group (T0001, T0024 and T0003). The ABAGYAN (T0009 and T0024) and STERNBERG (T0017 and T0003) groups also had very good success with some targets. It is also notable that particular modeling programs do better in the hands of their authors than when they are used by others, e.g. MODELLER [6,24].

In reviewing the six targets with multiple predictions available for assessment, there were no clear advantages for either database methods or *ab initio* methods. One point to note is that even for groups using the same program and method (e.g. MODELLER by FORSTER, ŠALI, SUTCLIFFE and WOLYNES), there are substantial differences in loop positioning, presumably due to differences in sequence alignment.

Fold recognition

Given a new sequence with no obvious homologue in the structure database identifiable by sequence alignment algorithms, there are two possible strategies for producing a model: fold recognition and *ab initio* modeling. In this section, we review the results for searching the structure database for a fold that is compatible with the new sequence. In the next section, we review *ab initio* methods that do not proceed from a known structure in the PDB.

The targets made available for fold recognition and *ab initio* modeling are listed in Table 3. The targets are divided by whether or not there were folds in the PDB that were identifiably similar with the DALI algorithm of Holm and Sander [25]. Some of the structural overlaps of PDB chains and CASP2 targets (Table 3a) make up only a small portion of the target protein, as low as 10–15% in some cases (T0002/1PSD-A and T0022/1MIO-B). We have listed the structures identified by DALI as having the lowest rmsd between the target and PDB entry Cα atoms and those having the largest overlap length with reasonably good rmsd scores. There were 17 targets whose structures were evaluated by the assessors, at least five of which appear to be new folds (Table 3b). The remaining 12 proteins (Table 3a) have domains or subdomains that can be matched to domains or subdomains of chains in the PDB.

The process of fold recognition is usually described as 'threading' the sequence of a new protein with unknown structure through any number of candidate folds derived from the structure database. As an example, in Figure 1 we show the S1 motif of nucleotidyl transferase (target T0004) [26] next to an image of 1CSP [27], which was identified as a related fold by several groups in the fold recognition portion of CASP2. The threading process involves creating an alignment of the target to the template fold sequences that optimizes some function of sequence identity, sequence variability in multiple aligned sequences of related proteins, hydrophobic burial and hydrophobic contacts, similarity of secondary structure propensities between target and template, and accessibilities. There are a variety of ways of optimizing these functions, i.e. ways of creating the sequence/sequence alignment.

Fold recognition depends on the size and quality of the fold library searched for structure matches. In a few cases, predictors used libraries of fewer than 300 folds (FINKELSTEIN, 146; SANEJOUAND, 188; LATHROP, 222; and DUBCHAK, 254 chains). Most predictors used 500–1000 PDB chains. Several groups used over 1000 chains, which must comprise almost all chains in the PDB (KARPLUS, 1114; HUBBARD, 1364; SIPPL, 1401; TORDA, 1464; ABAGYAN, 1557; COULSON, 4507; and MURZIN, 6359 chains). Computationally intensive methods must use smaller libraries, while faster methods can afford the redundancies of the larger libraries seen here.

Several groups made use of automated servers for at least part of their prediction procedure. BLAST and other

Table 3

Fold recognition and *ab initio* modeling targets.

(a) Targets with similar folds in the PDB.

Target ID	No. residues	Resolution	Name of protein	Species	Most similar PDB chains*	No. residues of overlap	Overlap rmsd for most similar chain	% Seq. ID
T0002	514	n.a.	Threonine deaminase	*Escherichia coli*	1PSD-A	70	2.24	9
					1WSY-B	303	2.57	20
T0004	84	NMR	Nucleotidyl transferase, S1 motif	*Escherichia coli*	1CSP	61	2.19	21
					1MJC	64	2.38	23
T0010	456	2.8 Å	Bactericidal/permeablility-increasing protein	*Homo sapiens*	1YTB-A	59	2.58	7
					2BBK-H	47	2.86	17
T0012	107	3.2 Å	Proregion of procaricain	*Carica papaya*	1PRI	55	10.35	5
					1GPB	61	13.11	11
T0014	252	2.5 Å	3-Dehydroquinase	*Salmonella typhimurium*	1DBS	71	2.81	7
					1UBS-A	192	3.04	15
T0016	312	2.0 Å	Peridinin chlorophyll protein	*Amphidinium carterae*	1CPC-A	69	4.26	9
T0020	320	1.9 Å	Ferrochelatase	*Bacillus subtilis*	3CHY	71	2.71	11
					8ABP	205	4.96	8
T0022	591	2.5 Å	L-Fucose-isomerase	*Escherichia coli*	1MIO-B	74	2.55	11
					1ENY	133	3.46	7
T0031	242	n.a.	Exfoliative toxin A	*Staphyloccous aureus*	3EST	185	2.31	14
					1TON	180	2.33	18
T0037	109	2.0 Å	Calponin homology domain of β-spectrin	*Homo sapiens*	1WHT-A	58	2.86	7
					2MHR	60	3.64	15
T0038	152	NMR	CBDN1, fructose-1,6-bisphosphatase	*Cellulomonas fimi*	1SLT-B	74	2.62	14
					2AYH	126	3.26	13
T0042	78	NMR	NK-lysin	Pig	1GPB	48	2.60	4

(b) Targets without similar folds in the PDB.

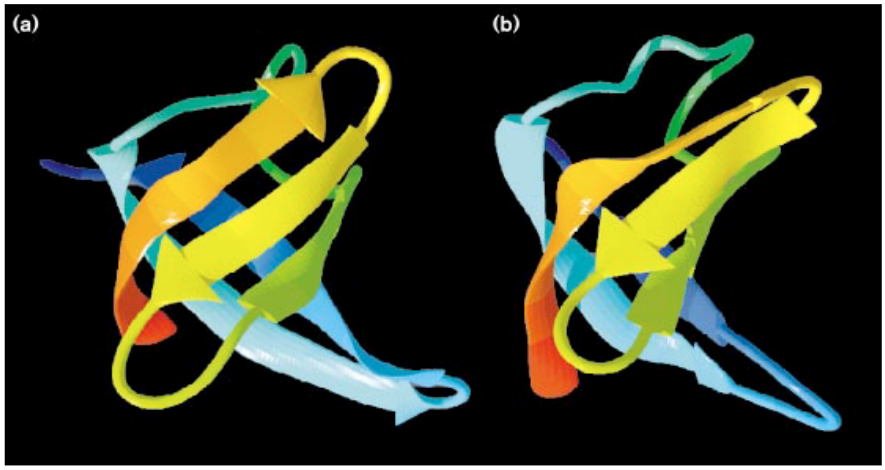
Target ID	No. residues	Resolution	Name of protein	Species
T0005	268	2.1 Å	γ-Fibrinogen C terminus	*Homo sapiens*
T0008	29	2.1 Å	*De novo* designed peptide	
T0011	220	1.8 Å	Hsp-90 N-terminal domain	*Saccharomyces cerevisiae*
T0030	66	NMR	Domain 1 of protein g3	Filamentous phage fd
T0032	98	2.2 Å	β-Cryptogein	*Phytophthora cryptogea*

*Chains listed here are the most similar chains of all CASP2 submitted predictions for each target, as judged by the DALI program [25].
n.a., not available.

sequence search methods were used to find related protein sequences for use in threading and secondary structure prediction methods that rely on multiple alignments. Automated secondary structure prediction methods were also used extensively, including the PHD server of Rost and Sander [28,29], the SSPRED server [30] and SOPMA [31]. PHD was also used for the calculation of accessibilities (STERNBERG and KARPLUS), secondary structure propensities (BARTON, COULSON, KARPLUS, MURZIN, SANDER and STERNBERG), sequence searches (BARTON, COULSON and HUBBARD), and fold recognition (ROST, PAZOS and MOULT).

Several groups used statistical approaches such as hidden Markov models, including the KARPLUS, HUBBARD and MUNSON groups. On the other hand, COULSON relied on existing tools (Smith–Waterman pairwise comparisons, Dayhoff scoring tables, the PHD program from Rost and Sander and the DSSP database by Holm and Sander) which were combined and optimized to produce a novel

Figure 1



Close fold similarity between (a) T0004 and (b) 1CSP. Ribbon representations of the S1 motif in nucleotidyl transferase [26] and the cold shock protein nucleotide-binding domain [27] based on secondary structure assigned by STRIDE [63] for T0004 and by the crystallographers for 1CSP. The two structures share a common folding topology known as the OB-fold (oligosaccharide/oligonucleotide-binding fold) [36] which was well recognized by the fold recognition methods. The chains are rainbow-colored from red (N termini) to blue (C termini). Ribbons were generated with MIDAS [64] and rendered with RAYSHADE [65].

package (Secondary Structure Alignment program, or SSS_ALIGN). JONES ranked folds with an improved pairwise potential function that he derived with Thornton [32,33]. Other approaches which depended on descriptors of local fold environments included PROFILES [34] (EISENBERG), training of a neural network (DUBCHAK) and the use of heuristics (LENGAUER).

Goodness of fit between sequence and fold may also be estimated from the energy of the sequence when threaded through a fold. SIPPL ranks alternative folds using potentials-of-mean-force [35] (PROFIT). OLSON derives his potentials from Boltzmann-like statistics based on work from Finkelstein (Institute of Protein Research, Moscow). These algorithmic methods are in contrast to the MURZIN group approach, which is rooted in the intimate knowledge of protein fold and function the authors derived from building the SCOP database [36].

The assessments were compiled by Steve Bryant (National Library of Medicine) and Krzysztof Fidelis (Lawrence Livermore National Laboratory) and were discussed at the conference by Michael Levitt (Stanford). We reviewed the assessments by taking the best submission for each target by each group (of the targets in Table 3a) and determining the number of targets for which each group had a threading specificity greater than 10%. The threading specificity was defined by the assessors as follows: each predictor team could place a certain portion of a bet on as many PDB chains as they wished, with the total bet adding up to 1.0. The assessors used the structure/structure comparisons to find all similar folds in the database to the target protein and gave these similarity scores with a score of 1.0 for the most similar chain and smaller scores for less similar chains (as defined by the structure/structure comparison method). The threading specificity was then computed as the following sum: threading specificity = $100 \times [\sum_i \text{Bet}(i) \times \text{Similarity}(i)]$, where the sum was over all chains in the submitted prediction. A threading specificity of 10% could mean, for instance, that a group put 10% of its bet on the correct protein and the rest of its bet on incorrect proteins, meaning they successfully narrowed the search for the correct fold to one of 10 chains. It could also mean that they placed all of their bet on one protein which was only 10% as similar as the best protein they could have picked from their dataset.

With this simple criterion, several groups performed quite well. The MOULT and THOMAS groups made predictions for only two of the targets in Table 3a, and were 2 for 2, or 100%. MURZIN was successful at this level for 4 out of his 5 predictions. The next several groups and their number of correct predictions versus number predicted were: SANDER (6/8), EISENBERG (7/10), BRYANT (6/9), ALEXANDROV (6/9), COULSON (7/11), TAYLOR (3/5), HUBBARD (5/9), and JONES (6/11). Another 13 groups had success rates between 25% and 50%, and the remaining eight groups were below 25%. What was very impressive was that these successes came in the guise of very diverse techniques.

Predictors were also required to provide sequence alignments for the target against the PDB folds on which they had placed non-zero bets. We looked at the alignment results by observing which groups produced predictions with 'alignment sensitivities' over certain thresholds. The alignment sensitivity was defined as the ratio of the number of correctly aligned pairs in the prediction to the number of aligned pairs in the structure/structure alignments. A small number of groups achieved alignment sensitivities over 50% on more than one target: ALEXANDROV, COULSON, EISENBERG, JONES, MURZIN, PAZOS, SANDER, and SIPPL. Clearly, this is an important goal and there is room for improvement.

One outcome of the fold recognition CASP2 experiment that is noteworthy is that this kind of prediction method played a role in both the comparative modeling and *ab initio* prediction sections. In at least two instances (ABAGYAN and MOULT), predictors used their recognized folds to produce atomic coordinates for a comparative model for their targets. This kind of model can be produced when the identified fold in the database is clearly homologous (i.e. similar from an evolutionary point of view). Fold recognition is therefore used to substantial advantage as a tool for remote homologue identification.

It is not clear at this point how often folds that are similar because of convergent evolution (without a common ancestor) are identified by threading techniques. Moreover, such identifications may occur in fragments of two or three secondary structural units with similar tertiary orientations to one another. Such information may be useful in building more complete models, but is probably of lower accuracy than when the folds are evolutionarily related. In this form, fold (or more accurately subfold) recognition methods may become part of the tools used in *ab initio* structure prediction, as indeed occurred in that section of the meeting (described below).

At CASP3, perhaps there should be no fold recognition category *per se*. Rather, in cases where there is no obvious homologue prior to entering a target onto the website, it may be left up to the predictors to submit whatever kind of prediction they are able to: the identification of a fold in the database, a predicted secondary structure, segment contacts between secondary structural units, or a full model with atomic coordinates (see below). If the predictors are complete in their description of what methods they used to make their models, the assessors can then be the judges of what techniques contributed most successfully to which targets.

Ab initio structure prediction

In the '*ab initio*' category of the CASP experiment, the data analysis and assessment teams led by Tim Hubbard and Arthur Lesk (MRC, Cambridge) were faced with challenges almost as demanding as the structure prediction problem itself. Long before the meeting, the design of a unified submission format posed particular problems. As the least well defined among the CASP categories, this section was open to all approaches that do not presume the existence of a similar folding topology to the target fold in the current PDB. In 1994, two rather different types of methods had been used in CASP1 *ab initio* predictions [37]. Most of the predictions had been submitted by groups using force field based or 'simulation' methods, where computational search algorithms attempt to find the conformation of the target protein molecule at the global minimum of free energy. The second type of submissions had come from groups extracting information from multiply aligned sequences of homologous proteins (of unknown structure) to provide clues as to the folded structure of the target protein. These methods will be referred to as knowledge-based approaches in the following. Tertiary structure predictions through knowledge-based approaches are usually made using manual or computer-assisted procedures that allow the incorporation of functional knowledge or hypotheses in the predictions. CASP1 had shown that the two different types of methods are distinct in the size range of suitable target proteins, or fragments, and in the level of resolution that can be expected from the resulting models. Further, *ab initio* methods are the farthest away from routinely producing tertiary structure predictions that would resemble the true structures of the target proteins.

Hubbard and Lesk recognized that the CASP1 findings would not allow fair assessment of the *ab initio* approaches if they designed a single unified submission format for CASP2. Instead, they allowed submissions at different levels of abstraction of protein structure using three main submission formats: secondary structure, low-resolution tertiary structure (encoded in a segment contact description of the folding topology developed by Lesk [38]) and high-resolution tertiary structure (in atomic coordinate format). A suite of programs was developed before the prediction meeting to calculate the lower resolution data (secondary structures, segment contacts, etc.) from atomic coordinates.

Still, the major challenge came with the evaluation of the *ab initio* submissions. The prediction challenges varied in many aspects (size of the target proteins, availability of homologous sequences, etc.) as demonstrated by the list of *ab initio*/fold recognition targets in Table 3. Furthermore, some protein folds presented more demanding prediction challenges than others. Thus, a comparative assessment of different prediction methods can be derived only from submitted predictions for the same target proteins. However, most prediction methods are not suited to make predictions for all target proteins. Thus, with the exception of the secondary structure level (submission format 1.1. [39]), the assessors had to rely mostly on their expert judgment, in contrast to the more quantitative assessment presented by the assessors in the other categories. Due to the large diversity in the methods, varying degrees of target difficulty and model resolution, there were no clear winners in the *ab initio* evaluation. Instead, the assessors presented an interesting mixture of predictions and predictor assessments by groups who had caught their attention with any particular aspect of their models, in some instances more local and in others more global structural features.

Secondary structure predictions (format 1.1 [39])

The use of numerical evaluation parameters for assessing the quality of predicted secondary structures as putative

starting points in tertiary structure modeling is a well known subject of debate in the field [37,40]. Thanks to the large numbers of assessable predictions at this level (a total of 169 submissions made explicitly in format 1.1 by 18 different teams and 7 automated servers, see Table 4), we can nevertheless derive valuable comparative conclusions about the performance of some of the methods used in CASP2 *ab initio* submissions.

Of central importance for the users of predictive tools in experimental laboratories, a comparison of internet servers identified the three most reliable automated secondary structure prediction tools that are currently accessible through the Web (as estimated by three-state percentage score (*Q3*) and segment overlap (*Sov*) [41] performance): PREDICTPROTEIN [29] (a profile-based neural network method using multiple sequence information, known as the PHD program, by Rost and Sander (EMBL, Heidelberg) [28]; DSC [42] (a heuristic-based method using multiple sequence information and involving linear discriminant statistics in prediction refinement, by King and Sternberg (Imperial Cancer Research Fund, London) [43]; and NNSSP (an improved nearest neighbor method [44] using multiple sequence information by Salamov and Solovyev (Baylor College of Medicine, Houston) [45].

The good news is that these fully automated systems were only marginally inferior to the best expert predictions involving manual intervention (including those generated by the creators of the servers by manual refinement of the automated outputs). The bad news is that, as might to be expected from theoretical considerations [46], it seems unlikely that secondary structure predictions will become generally error-proof in the future. Consequently, useful tertiary structure assembly approaches will have to account for the possibility of occasional errors in secondary structure predictions that are used as a starting point for tertiary structural models.

An interesting demonstration of a central purpose of the CASP experiments came through the assessment of the secondary structure predictions derived from the 3D models generated by LINUS, a highly publicized automated force field based prediction system by Srinivasan and Rose (Johns Hopkins University, Baltimore) [47]. Although the system could be used on only a few selected targets, it was apparent that LINUS's promising secondary structure prediction performance tested on known protein structures could not be reproduced in a blind prediction setting, at least not with the current version of the program.

Tertiary structure predictions

The goal of the CASP experiments reaches beyond secondary structure to the prediction of tertiary structures. Emboldened by some successes in CASP1, some of the prediction teams attempted the assembly of secondary structure segments into 3D protein models (Table 4). Adding up all submissions made explicitly in one of the tertiary structure prediction formats (*2.1*, *2.2*, or 3.1, see below) against experimental structures that were available by the time of the meeting, a total of 73 submissions described the predictions by 19 different teams (Table 4).

Assembled tertiary structure models, however, presented the assessors with another problem. Although *ab initio* methods do not presume the existence of a similar fold among the currently known structures, predictor teams following semi-automated or manual approaches often detected (or hypothesized) structural similarities with known folds at later stages of the prediction process. Lesk and Hubbard decided to consider such models separately, together with other submissions where manually built models were based on fold recognition predictions by automated methods. Both of the final sets of predictions, the true *ab initio* and the known fold based *ab initio* predictions, were analyzed with the same automated evaluation tools and are available through the Web pages of the Protein Structure Prediction Center at Lawrence Livermore National Laboratory [4]. Where predictions were based on correctly predicted fold similarities, these evaluation numbers provide us with an important 'internal standard' for 3D models built at low-resolution levels.

An interesting illustration for the difficulties in defining the boundaries between the fold recognition and the *ab initio* categories is provided through the prediction submissions by the MURZIN group. While these predictions were generated through an approach most similar to the knowledge-based method type categorized as *ab initio* in CASP1, their most remarkable overall accuracy allowed them to be evaluated with the higher standards of the fold recognition category.

However, not all CASP2 folds had precedents in the PDB. Compared with CASP1, where common superfolds (α/β-barrels, Greek key β-sandwiches, etc.) were overrepresented among the targets [37], a larger number of unprecedented folding topologies, some with rather uncommon features, were seen in the CASP2 target proteins (Table 3b; Fig. *2*), and only few of the known topologies had been designated superfolds prior to the CASP2 prediction season. While a closer analysis of the less commonly observed structural features in the new folds will be very useful for refining prediction heuristics, the sample of folds to predict clearly made CASP2 a challenge of higher difficulty than CASP1.

Topology/segment contact predictions (formats 2.1/2.2 [39])

A new segment contact format was developed by Lesk [38] to describe protein folds at very low levels of resolution, similar to that of wire models. While segment contact descriptions can be derived from atomic coordinates, it was

Table 4

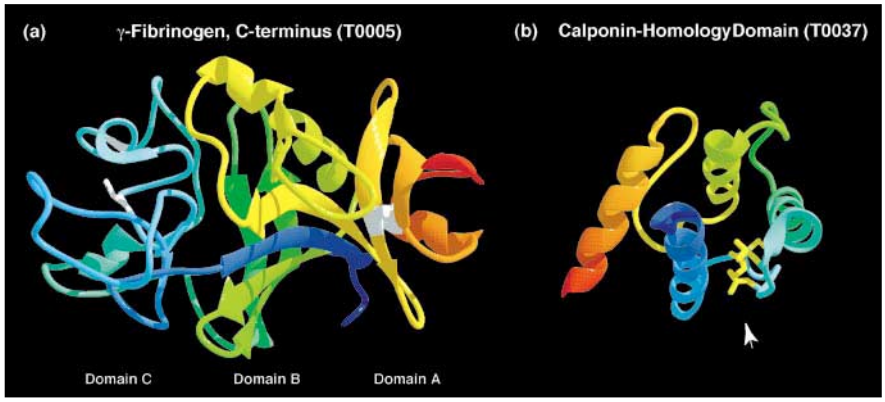
Assessable *ab initio* predictions sorted by participating group and format of submission.

Prediction teams (named after the group leaders unless indicated otherwise)	Entered targets in format 1.1 (explicit only)	Explicit 2.1 entries [total (FR)]*	Explicit 2.2 submissions [total (FR)]*	3.1 submissions [total (FR)]*
Force field based methods:				
ABAGYAN	6	–	–	6 (5)
AVBELJ	–	–	–	3
EISENBERG	–	–	–	1
MOULT	–	–	–	6 (2)
OSGUTHORPE	–	–	–	2
ROSE	3	–	–	1
Total	9	–	–	19† (7)
Other automated methods:				
BAKER [local structure elements from sequence]	–	–	–	9
JONES [assembly from short threading fragments]	–	–	–	1
LENGAUER [combinatorial algorithms, graph theory]	4	4	–	–
MARSHALL [contact matrix method]	–	–	–	2
SMITH [statistics method, folding class prediction]	3	1	–	1
TAYLOR [recognition of self-generated folds]	–	–	–	3 (1)
VALENCIA [contact prediction]	4	4	3	–
Total	11	9	3	16 (1)
Predictions based on automated fold recognition:				
HUBBARD	5	1 (0–1)	–	1 (1)
STERNBERG	16	–	–	1 (1)
Total	21	1 (0–1)	–	2 (2)
Knowledge-based methods:				
BAZAN	2	2 (1–2)	–	1 (1)
COHEN‡	6	6 (2)	–	4 (2)
FINKELSTEIN	5	1	–	–
MURZIN	1	–	–	9 (7)
Total	14	9 (3–4)	–	14 (10)
Secondary structure prediction only:				
BENNER	2	–	–	–
GOLDSTEIN	8	–	–	–
JAAP	15§	–	–	–
MUNSON	10	–	–	–
ROST	16	–	–	–
SHESTOPALOV	6	–	–	–
SOLOVYEV	15	–	–	–
SERVER-COMPARISON: 7 automated internet servers (listing on http://september.llnl.gov::8000/AbInitioEvaluation)	6 each	–	–	–
Total	114§	–	–	–

*Total number of predictions in each format that were submitted by the respective teams, where experimental structures were available at the meeting; numbers in parentheses indicate how many predictions were based on fold recognition (FR) hypotheses. †One prediction for T0009, a comparative modeling (not a fold recognition/*ab initio*) target, is included. ‡Collaboration between the Benner and Cohen groups. §Five secondary structure predictions for comparative modeling (not fold recognition/*ab initio*) targets are not listed.

also possible to submit predicted contacts and approximate relative orientations between predicted secondary structure elements alone. Thus, submissions in this format were most suitable for predictions generated by knowledge-based methods, where atomic coordinate models might not have been generated. Accordingly, topology descriptions using the new code were mostly the domain of this type of method (Table 4), although only a few predictions were entered explicitly in 2.1 or 2.2 formats in CASP2. Lesk's description code is likely to become popular in future

Figure 2



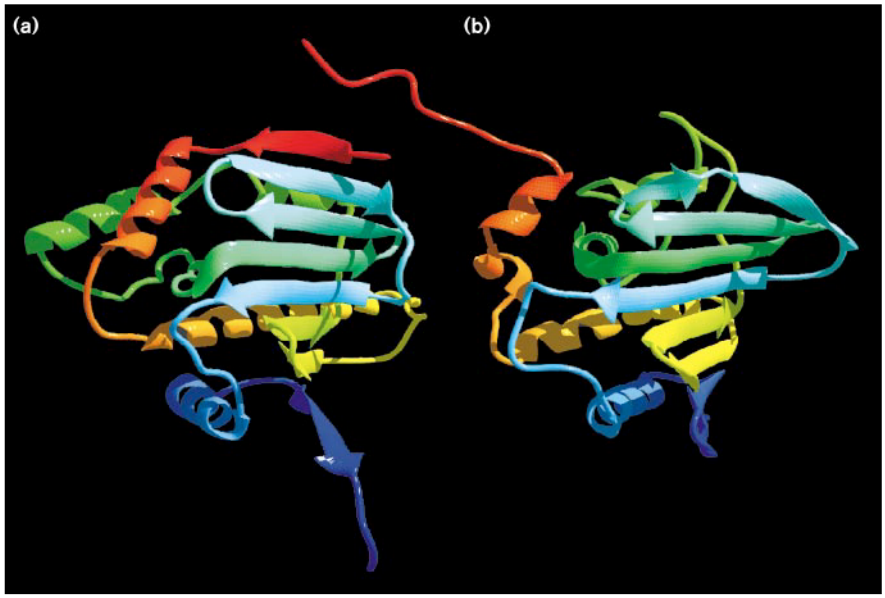
Some of the interesting new folds among the CASP2 targets. Ribbon representations of (a) T0005, the γ-fibrinogen C-terminal fragment [66] and (b) T0037, the calponin homology domain [67], based on secondary structure assignments by STRIDE [63]. The chains are rainbow-colored from red (N termini) to blue (C termini). The C terminus of γ-fibrinogen folds in three domains in an unprecedented topology with strongly twisted and curled β-sheet structure. Two disulfide bridges in domains A and C are shown in white. Domain C is unusual in that it almost lacks regular secondary structure (α-helices or β-strands). The folding topology of the calponin homology domain could be described as an all-parallel four-helix bundle, an unusual topology for a stand-alone structure. Alternatively, if the N-terminal helix is separated, the structure can be viewed as an up-up-(down)-up bundle, where the third helix has been replaced by a coil structure contributing to the hydrophobic core of the bundle (yellow sidechains). Ribbons were generated with MIDAS [64] and rendered with RAYSHADE [65].

CASP events as it allows the submission and evaluation of predicted arrangements of core fragments of the target structures without overemphasizing inserted loop segments that require much more sophisticated modeling procedures (i.e. those of comparative modeling).

In their assessment at the topology description level, Lesk and Hubbard used the number of correctly predicted contacts between predicted secondary structure elements as a principal evaluation measure. With the exception of those predictions where models were based on correctly recognized folds, the performance of knowledge-based methods on medium-sized proteins seemed weak, but increased where long-range distance constraints (including information regarding domain boundaries) could be identified, and mostly reflected the recognition of general folding class features (e.g. mostly parallel β-sheet versus anti-parallel β-sheet structures).

An interesting example of knowledge-based fold recognition was provided by the N-terminal fragment of heat shock protein 90 (T0011; LH Pearl, personal communication). Both the BAZAN and COHEN teams identified putative active site residues through multiple sequence analysis, while the catalytic activity of the target protein was debated in the biochemical literature. Together with

Figure 3



Partial fold similarity between (a) T0011 and (b) the DNA gyrase N-terminal domain. Ribbon representations of the heat shock protein 90 N terminus (LH Pearl, personal communication) and the ATPase domain of DNA gyrase B [48] based on secondary structure assigned by STRIDE [63]. Although the gyrase structure was not available through the PDB at the time of the contest, the similarity between the two structures was recognized by two groups in the *ab initio* category using knowledge-based methods (see text). The chains are rainbow-colored from red (N termini) to blue (C termini). Ribbons were generated with MIDAS [64] and rendered with RAYSHADE [65].

supersecondary structure predictions for the fragment, reconstructions of a putative ATPase active site led both teams to recognize the topological similarity with the N-terminal domain of the published structure of DNA gyrase [48], which was not available through the PDB at the time (Fig. 3). The observation that two independent groups reached the same conclusion is particularly important in this case. It indicates that the problem was indeed approached systematically, making it probable that similar procedures can be implemented in automated systems in the future.

For those targets with new folds (Table 3b; Fig. 2), however, CASP2 mostly confirmed the helplessness of the predictors towards such problems pointed out in the CASP1 assessment [37]. Nevertheless, the topology description code (*2.1* and *2.2.* formats) and the corresponding assessment parameters clearly define a path of gradual improvement in topology prediction. While the number of contacts between secondary structure elements was used to assess the CASP2 results, the goals for CASP3 should also include the prediction of the approximate angles between such elements, and possibly the identities of the corresponding contact residues (format *2.2*).

Coordinate models (format 3.1 [39])

The highest level of resolution in model assessment was possible with the entries submitted in atomic coordinate format. Most coordinate submissions came from groups using force field based and similar automated methods, with the length of the predicted fragments varying from 6 to 108 amino acid residues for this type of method. In addition, some of the predictions generated through the knowledge-based approaches (see Table 4) were submitted as explicit coordinate files. As expected, based on the different types of methods, the automated evaluation of the number of equivalent Cα atoms in the longest superimposable fragment revealed generally shorter fragment matches for the force-field methods (up to three secondary structural elements in length) than for their knowledge-based counterparts, but at higher levels of accuracy (i.e. lower rmsd). In particular, some progress was noted in the longer, more exact, matches than those achieved in CASP1, especially in the regions that lie between segments of regular secondary structure.

An interesting novel approach for predicting local tertiary structure was presented by the BAKER group. By identifying local structure elements (e.g. internal helices or specific types of turns) from sequence patterns and using them as starting points in their automated assembly in tertiary structure models, the BAKER approach differs from the force field based and from the knowledge-based methods. At CASP2, an unrefined version of the algorithm produced the most accurately matched fragment over the length of three consecutive helical segments (residues 19–48 in L-fucose isomerase, T0022). While the global fold predictions using the same method have not been as successful, the potential of this innovative approach has been noted by many experts in the field.

In the prediction of longer segments, it was the three models for NK-lysin (T0042) by the JONES group that yielded the longest superimposable segments in predictions that were not based on a fold recognition prediction. The JONES models scored from 39 to 53 equivalent Cα atoms over a superimposable fragment between 68 and 75 residues in length, reflecting the topological correspondence of these models with the actual NK-lysin structure. This success was, in part, enabled by the recognition of a kink in a long helical segment that had escaped most other approaches.

Finally, the importance of considering homo-oligomeric structure in structure prediction was stressed with the structure of T0008, a 29-residue designed protein [49]. Although the target description had hinted at it, only the ABAGYAN team had worked with a multimeric fold, unfortunately with a dimer instead of a trimer.

Thus, the medium-resolution CASP2 successes in the true *ab initio* section are still restricted to fragment matches in individual models by individual prediction groups and on individual target proteins, leaving us with the feeling that there is much to be learned before we can make accurate predictions for larger protein structures with these computational techniques. Also, accurately matched fragments generally involved α-helical segments of the proteins, while prediction of the more irregular β-strand conformations still seems to be out of reach at the coordinate level. CASP2 has shown improved performance towards meeting the challenges of *ab initio* structure prediction and suggests that the combination with innovative methods like the approach taken by the BAKER team could prove beneficial in medium-resolution prediction of protein fragments from single sequences.

Redefining the goals of ab initio structure predictions

With the improved performance of the automated fold recognition methods in CASP2 and the rapidly increasing number of experimental structures, an important question emerges: will *ab initio* prediction efforts soon become obsolete? To answer this question adequately, we must consider not only the immediate goal of reproducing experimental structures through theoretical methods, but also what we can gain from the *ab initio* CASP experiment and from further development of the different types of *ab initio* approaches.

For example, in a time when sequences from genome projects are providing us with large sets of homologous sequences for any protein of functional interest, what use

is there in attempting to predict the tertiary structure of a protein with force field based methods that disregard the benefits of multiple sequence information? The answer lies in the observation that the tertiary structures of homologous proteins are generally similar in their structural core segments but can differ drastically in other regions. Structural variation among homologous proteins in non-core regions does not generally imply that these regions are functionally unimportant and neutral towards evolutionary selection pressure. Instead, structural variation sometimes emerges where evolution has worked towards functional adaptation of individual members of a protein family, e.g. by engineering a binding site for a different cofactor. The only readily accessible information for predicting the structure in these regions is contained in the individual target sequence, and this demonstrates the need for single-sequence prediction methods.

Furthermore, we might also ask whether it is useful to put together multiple sequence alignments and wade through the biochemical literature following knowledge-based approaches, when notable CASP2 successes for medium-sized proteins were mostly models based on manual sequence/structure alignments with known structures. Sequence/structure alignments of similar quality might soon become available through fully automated fold recognition methods, drastically diminishing the time expense for predictions of this kind. Here, the point is best illustrated by the exceptional results achieved by the MURZIN team with their manual tertiary structure assemblies. While it is true that these results could not, at this time, be reproduced by less knowledgeable scientists in the field (AG Murzin has central responsibilities in the manual fold classification for the SCOP database [36]), the fact that the team was able to identify sequence/structure relationships that allowed them to outperform automated methods is crucial. It indicates that there is an open range of heuristics to be discovered in this area, most probably through closer inspection and understanding of the essential sequence requirements for particular folding topologies. Additionally, the power of multiple sequence analysis lies in the detection of functional aspects of the target proteins that can be paired with the structural aspects in the development of new heuristics for tertiary structure prediction of all categories.

Therefore, the goals in the *ab initio* category should be less target oriented, and more method oriented, than in the other categories. The emphasis should be to isolate individual steps in the prediction procedures that can provide information (structural and functional) complementary to other prediction methods and that can increase our understanding of structural and functional requirements of proteins. Understanding of this kind is needed not only in structure prediction challenges, but also in protein design and the recognition of structure/function relationships.

Docking

A new category of predictions was added to CASP2: the docking of molecules to proteins. Participants were asked to predict the modes of association between ligands and proteins. Eight targets were available for predicting small-molecule ligand docking and one target was provided for predicting protein–protein docking (Table 5). Scott Dixon of Smith–Kline Beecham (King of Prussia, PA) was the assessor for this category.

For small molecule ligand–protein complexes, the uncomplexed structures of proteins were either available from the PDB or provided by the experimentalists who were solving the structures of the complexes. The predictors were given the chemical structure of the ligands, but were required to predict their conformations. For protein–protein docking, the challenge was to predict the binding mode for a very large complex, given the prior knowledge of the uncomplexed structures of the two proteins. The main criterion for evaluation was rmsd of the ligand heavy atoms from the experimental structures. In the case of the protein ligand, only atoms within 5 Å of the receptor protein were used to calculate rmsd. The percentage of correct contacts between a ligand and a receptor was also considered. Contacts were defined as those receptor–ligand atom pairs with a distance less than the sum of their van der Waals' distances plus 25%.

Twelve groups [50–56] submitted a total of 43 predictions for the eight protein–ligand targets. Since the active sites were known for all proteins involved in small-molecule binding (trypsin accounted for four targets, elastase for two targets, concanavilin and fructose-1,6-bisphophatase for one each), these test cases were not particularly difficult. A successful model should include the correct conformation for the ligand as well as its binding mode. Overall, predictors fared relatively well in these tests, being able to predict for each target at least one binding mode close to experimental results. The best model for each target had rmsd from the experimental position of less than 5 Å. In particular, predictions for targets T0013 (complex of concanavalin A and methyl-D-arabinofuranoside) and T0034 (complex of trypsin and amiloride) were very close to experimental results. For most models, rmsds were less than 3.5 Å. The two best models found virtually the right answers for these two targets, with rmsds of 1.4 Å by the READ group (target T0013) and 0.6 Å by the MITCHELL group (target T0034).

Some difficulties were presented by two ligands of pancreatic trypsin that did not interact extensively with the protein, but rather stuck out from the binding pocket in the experimental structures. Of these two (T0035, pancreatic trypsin–SBA complex, and T0036, pancreatic trypsin–SBB complex), predictors did better against

Table 5

Docking targets.

Target ID	No. residues	Resolution	Name of protein	Ligand	Species
T0013	237	1.7 Å	Complex of concanavalin A	Methyl α-D-arabinofuranoside	*Canavalia ensiformis*
T0018	950	n.a.	Hemagglutinin	Fab	*Influenza* virus
T0033	231	n.a.	Trypsin	Pentamidine	*Bos taurus*
T0034	231	n.a.	Trypsin	Amiloride	*Bos taurus*
T0035	240	n.a.	Pancreatic elastase	SBA	*Sus scrofa*
T0036	240	n.a.	Pancreatic elastase	SBB	*Sus scrofa*
T0039	335	2.6 Å	Fructose-1,6-bisphosphatase	Aica-riboside phosphate	*Homo sapiens*
T0040	245	2.6 Å	Pancreatic trypsin	INH	*Bos taurus*
T0041	245	2.6 Å	Pancreatic trypsin	INI	*Bos taurus*

n.a., not available.

trypsin–SBB, because SBB forms a covalent link with trypsin, providing an additional constraint for modeling.

With a few exceptions, proteins were assumed to be rigid while ligands were either rigid or flexible. Docking was accomplished by Monte Carlo and other sampling methods. Ranking of generated models was based on interaction energy and/or shape complementarity (scoring functions). Solvent molecules were not considered for most models. Sampling the binding modes appeared to be a less difficult problem than developing more accurate scoring functions, which can ultimately help design stronger binding ligands.

A number of factors contributed to the relatively successful predictions: well-formed binding pockets were known to predictors; structures of protein–ligand complexes similar to the targets were available; and ligands were small and could adopt only limited conformations. For example, amiloride is small and highly conjugated, resulting in a nearly planar conformation. It is not surprising that predictors had most success against this target (T0034).

Predicting the mode of protein–protein associations (target T0018, hemagglutinin–Fab complex) proved to be a much more difficult problem. Ideally, a successful prediction should identify the protein–protein binding interface and the correct binding mode. Three groups were able to identify the binding interface. GRAMM [57], an algorithm developed by Vakser [57,58], performed an exhaustive search for the binding site using an empirical intermolecular energy function. With the consideration of biochemical knowledge and prior experimental results, the STERNBERG group [59–61] performed a limited grid-based search and evaluated the complexes with a molecular mechanics force field including a solvation term. This was followed by limited rigid-body minimization. The DELISI group used the program DOCK [62] to dock the two proteins with a molecular mechanics potential including empirical terms representing solvation and entropic distributions to rank binding orientations. All three groups maintained the conformations of the individual proteins.

All three of these approaches were able to correctly identify the binding interface between hemagglutinin and the Fab antibody fragment. The rmsds ranged from 10 to 20 Å. This is very encouraging considering the large size of the complex. However, predicting correct binding mode was a difficult task, and the percentages of correctly predicted protein contacts were low. Such results demonstrate that it is possible to model the gross features of a large protein–protein complex at low resolution, but it is still very difficult to model detailed molecular interactions.

Out of the four prediction categories presented at CASP2, the docking targets were by far the most straightforward. These methods are of substantial practical importance in structure-based drug design. As the results from predictions for both protein–ligand and protein–protein complexes suggest, most algorithms managed to sample potential binding orientations thoroughly. However, the challenge remains to determine how to evaluate these possible binding modes to identify the correct one and to rank ligands with different binding strengths against the same receptor. It is clear from this meeting that the current generation of docking algorithms have some distance to go before becoming a quantitative tool for predicting detailed molecular interactions.

Conclusions

The field of protein structure prediction has matured rapidly in the past few years, and with larger experimental databases and faster computers should continue to do so.

The CASP process initiated by John Moult has provided the field with something nearly universal in most scientific work, but absent until recently in structure prediction: situations where one does not know the answer before doing the experiment. The uncertainty, we think, has added to the excitement in making the predictions for those involved in the process. Bringing individuals involved in many different aspects of structure prediction together at one meeting has also been important in combining methods and bridging techniques and goals of structure prediction.

The major developments between CASP1 in 1994 and CASP2 in 1996 have been the establishment of unified submission formats for each of the categories and the creation of the first series of systematic assessment parameters for each of the categories. While some of these parameters may have to be revised for use in future CASP events, the CASP2 automated assessment has enforced the important discussion about evaluation criteria for model predictions.

CASP2 has motivated approximately 70 prediction teams worldwide to spend far more time than usual looking at their models and at known protein structures. In an era when enormous numbers of known structures in the PDB promote highly sophisticated analyses of structural aspects over the entire set, it has done many of us good to go back and spend a summer admiring and studying the perfection of individual examples of protein structures.

Acknowledgments

We would like to thank the CASP2 organizers, but especially Krzysztof Fidelis and Steve Bryant for providing information not yet available on the CASP2 and Lawrence Livermore Protein Structure Prediction Center websites. Also, a special debt is owed to those X-ray crystallographers and NMR spectroscopists who provided sequences and atomic coordinates of proteins under experimental study prior to publication. In particular, we thank M Bycroft, V Yee, K Dijnovic Carugo, DB Wigley and LH Pearl for making coordinates available for the illustrations before their release through the PDB. RL Dunbrack is an NIH Post-doctoral Fellow. DL Gerloff is a fellow of the Leukemia Society of America. O Lichtarge is an American Heart Association fellow.

References

1. Moult, J. The ProStar Webpage. Center for Advanced Research in Biotechnology, University of Maryland. World Wide Web URL: http://prostar.carb.nist.gov/
2. Fischer, D. & Eisenberg, D. The UCLA-DOE Benchmark To Assess The Performance Of Fold Recognition Methods. University of California Los Angeles. World Wide Web URL: http://www.mbi.ucla.edu/people/frsvr/frsvr.html
3. Hubbard, T., *et al*., & Bryant, S. Second Meeting on the Critical Assessment of Protein Structure Prediction. Center for Advanced Research in Biotechnology, University of Maryland. World Wide Web URL: http://iris4.carb.nist.gov/casp2/
4. Fidelis, K. Protein Structure Prediction Center. Biology and Biotechnology Research Program , Lawrence Livermore National Laboratory, Livermore, CA, USA. World Wide Web URL: http://PredictionCenter.llnl.gov/
5. Cardozo, T., Totrov, M. & Abagyan, R. (1995). Homology modeling by the internal coordinate mechanics (ICM) method. *Proteins* 23, 403–414.
6. Šali, A. & Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.
7. Taylor, W.R. & Orengo, C.A. (1989). Protein structure alignment. *J. Mol. Biol.* 208, 1–22.
8. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics* 8, 52–56.
9. Bruccoleri, R. E. & Karplus, M. (1987). Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* 26, 137–168.
10. Moult, J. & James, M.N.G. (1986). An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. *Proteins* 1, 146–163.
11. Aszódi, A., Gradwell, M.J. & Taylor, W.R. (1995). Global fold determination from a small number of distance restraints. *J. Mol. Biol.* 251, 308–326.
12. Harrison, R.W., Chatterjee, D. & Weber, I.T. (1995). Analysis of six protein structures predicted by comparative modeling techniques. *Proteins* 23, 463–471.
13. Ponder, J.W. & Richards, F.M. (1987). Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* 193, 775–792.
14. Abagyan, R. & Totrov, M. (1994). Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* 235, 983–1002.
15. McGregor, M.J., Islam, S.A. & Sternberg, M.J.E. (1987). Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* 198, 295–310.
16. Schrauber, H., Eisenhaber, F. & Argos, P. (1993). Rotamers: to be or not to be? An analysis of amino acid side-chain conformations in globular proteins. *J. Mol. Biol.* 230, 592–612.
17. Levitt, M. (1992). Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.* 226, 507–533.
18. Chinea, G., Padron, G., Hooft, R.W.W., Sander, C. & Vriend, G. (1995). The use of position-specific rotamers in model building by homology. *Proteins* 23, 415–421.
19. Dunbrack, R.L., Jr. & Karplus, M. (1993). Backbone-dependent rotamer library for proteins: application to sidechain prediction. *J. Mol. Biol.* 230, 543–571.
20. Dunbrack, R.L., Jr. & Karplus, M. (1994). Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat. Struct. Biol.* 1, 334–340.
21. Dunbrack, R.L., Jr. The backbone-dependent rotamer library webpage. University of California, San Francisco. World Wide Web URL: http://www.cmpharm.ucsf.edu/~dunbrack
22. Bower, M., Cohen, F.E. & Dunbrack, R.L., Jr. (1997). Homology modeling with a backbone-dependent rotamer library. *J. Mol. Biol.* in press.
23. Bower, M., Cohen, F.E. & Dunbrack, R.L., Jr. (1997). SCWRL: a program for building sidechains onto protein backbones. University of California San Francisco. World Wide Web URL: http://www.cmpharm.ucsf.edu/~bower/scwrl.html
24. Šali, A. MODELLER page at the Rockefeller University. Rockefeller University. World Wide Web URL: http://guitar.rockefeller.edu/modeller/modeller.html
25. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233, 123–138.
26. Bycroft, M., Hubbard, T.J.P., Proctor, M., Freund, S.M.V. & Murzin, A.G. (1997). The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell* 88, 235–242.
27. Schindelin, H., Marahiel, M.A. & Heinemann, U. (1993). Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature* 364, 164–168.
28. Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods Enzymol.* 266, 525–539.
29. Rost, B. & Sander, C. PredictProtein. European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. World Wide Web URL: http://www.embl.heidelberg.de/predictprotein/ppDoPred.html
30. Mehta, P., Heringa, J., Milpetz, F. & Argos, P. SSPRED: a service for secondary structure prediction of proteins. European Molecular Biology Laboratory, Heidelberg Germany. World Wide Web URL: http://www.embl-heidelberg.de/sspred/sspred_info.html
31. Geourjon, C., Deléage, G. & Blanchet, C. Protein Sequence Analysis at IBCP. Institute of Biology and Chemistry of Proteins, Lyon, France. World Wide Web URL: http://www.ibcp.fr/mailserver.html
32. Jones, D.T., Taylor, W.R. & Thornton, J.M. (1992). A new approach to protein fold recognition. *Nature* 358, 86–89.
33. Jones, D.T., Miller, R.T. & Thornton, J.M. (1995). Successful protein fold recognition by optimal sequence threading validated by rigorous blind testing. *Proteins* 23, 387–397.
34. Fischer, D. & Eisenberg, D. (1996). Protein fold recognition using

sequence-derived predictions. *Protein Sci.* 5, 947–955.
35. Floeckner, H., Braxenthaler, M., Lackner, P., Jaritz, M., Ortner, M. & Sippl, M. J. (1995). Progress in fold recognition. *Proteins* 23, 376–386.
36. Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540.
37. Defay, T. & Cohen, F.E. (1995). Evaluation of current techniques for *ab initio* structure prediction. *Proteins* 23, 431– 447.
38. Lesk, A.M. (1995). Systematic representation of protein folding patterns. *J. Mol. Graphics* 13, 159–164.
39. Hubbard, T. CASP2: formats for submission of *ab initio* predictions. Centre for Protein Engineering, Cambridge, United Kingdom. World Wide Web URL: http://iris4.carb.nist.gov/casp2/ab-submission.html
40. Jenny, T.F. & Benner, S.A. (1994). Evaluating predictions of secondary structure in proteins. *Biochem. Biophys. Res. Commun.* 200, 149–155.
41. Rost, B., Sander, C. & Schneider, R. (1994). Redefining the goals of structure prediction. *J. Mol. Biol.* 235, 13–26.
42. King, R.D. & Sternberg, M.J. DSC: discrimination of protein secondary structure class. Imperial Cancer Research Fund, London, United Kingdom. World Wide Web URL: http://bonsai.lif.icnet/bmm/dsc/dsc_read_align.html
43. King, R.D. & Sternberg, M.J. (1996). Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci.* 5, 2298–2310.
44. Yi, T.-M. & Lander, E.S. (1993). Protein secondary structure prediction using nearest-neighbor methods. *J. Mol. Biol.* 232, 1117–1129.
45. Salomov, A.A. & Solovyev, V.V. (1995). Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments. *J. Mol. Biol.* 247, 11–15.
46. Benner, S.A., Gerloff, D. & Chelvanayagam, G. (1995). The phospho-β-galactosidase and synaptotagmin predictions. *Proteins* 23, 446–453.
47. Srinivasan, R. & Rose, G.D. (1995). LINUS: a hierarchic procedure to predict the fold of a protein. *Proteins* 22, 81–99.
48. Wigley, D.B., Davies, G.J., Dodson, E.J., Maxwell, A. & Dodson, G. (1991). Crystal structure of an N-terminal fragment of the DNA gyrase B. *Nature* 351, 624–629.
49. Ogihara, N.L., Weiss, M.S., DeGrado, W.F. & Eisenberg, D. (1997). The crystal structure of the designed trimeric coiled coil coil-V(a)L(d): implications for engineering crystals and supramolecular assemblies. *Protein Sci.* 6, 80–88.
50. Rarey, M., Wefing, S. & Lengauer, T. (1996). Placement of medium-sized molecular fragments into active sites of proteins. *J. Cell Mol. Dynam.* 10, 41–54.
51. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 261, 470–489.
52. Sobolev, V., Wade, R.C., Vriend, G. & Edelman, M. (1996). Molecular docking using surface complementarity. *Proteins* 25, 120–129.
53. Vajda, S., Weng, Z., Rosenfeld, R. & DeLisi, C. (1994). Effect of conformational flexibility and solvation on receptor–ligand binding free energies. *Biochemistry* 33, 13977–13988.
54. Weng, Z., Vajda, S. & Delisi, C. (1996). Prediction of complexes using empirical free energy functions. *Protein Sci.* 5, 614–626.
55. Xu, L.Z., Weber, I.T., Harrison, R.W., Gidh-Jain, M. & Pilkis, S.J. (1995). Sugar specificity of human beta-cell glucokinase – correlation of molecular models with kinetic measurements. *Biochemistry* 34, 6083–6092.
56. Weber, I.T. & Harrison, R.W. (1996). Molecular mechanics calculations on HIV-1 protease with peptide substrates correlate with experimental data. *Protein Eng.* 9, 679–690.
57. Vakser, I.A. GRAMM v1.03: global range molecular matching. Rockefeller University, New York, NY. World Wide Web URL: http://guitar.rockefeller.edu/gramm
58. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C. & Vakser, I.A. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA* 89, 2195–2199.
59. Gigant, B., Fleury, D., Bizebard, T., Skehel, J.J. & Knossow, M. (1995). Crystallization and preliminary X-ray diffraction studies of complexes between an influenza hemagglutinin and Fab fragments of two different monoclonal antibodies. *Proteins* 23, 249–275.
60. MacCallum, R.M., Martin, A.C.R., Thornton, J.M. (1996). Antibody–antigen interactions: contact analysis and binding site topography. *J. Mol. Biol.* 262, 732–745.
61. Koehl, P. & Delarue, M. (1994). Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* 239, 249–275.
62. Kuntz, I.D. (1992). Structure-based strategies for drug design and discovery. *Science* 257, 1078–1082.
63. Frishman, D. & Argos, P. STRIDE: protein secondary structure assignment from atomic coordinates. European Molecular Biology Laboratory. World Wide Web URL: http://www.embl-heidelberg.de/argos/stride/stride.html
64. Ferrin, T.E., Huang, C.R., Jarvis, L.E. & Langridge, R. (1988). The MIDAS display system. *J. Mol. Graphics* 6, 13–21.
65. Kolb, C.E. General rayshade information. Stanford University. World Wide Web URL: http://www-graphics.stanford.edu/~cek/rayshade/info.html
66. Yee, V.C., *et al.*, & Teller, D.C. (1997). Crystal structure of a 30 kDa C-terminal fragment from the γ chain of human fibrinogen. *Structure* 5, 125–138.
67. Dijnovic Carugo, K., Banuelos, S. & Saraste, M. (1997). Crystal structure of a calponin homology domain. *Nat. Struct. Biol.* 4, 175–179.